数理解析研究所講究録
623巻 1987年 1-12

経験分布関数に基づいた適合度検定統計量の漸近的性質について

統計数理研究所　安芸重雄（Sigeo Aki）

1. 序

Kolmogorov-Smirnov, Cramér-von Mises, Anderson-Darling 等の経験分布関数に基づいた適合度検定統計量の漸近分布を求める時，基本的な事は，一様分布からのランダム・サンプルに基づく経験過程 $\sqrt{n}(\Gamma_n(t)-t)$ が Brownian bridge $\{\beta(t): 0\le t\le 1\}$ へ弱収束するという事実である．このことは Kolmogorov-Smirnov 統計量を除けば Mises 汎関数或いは Symmetric statistics という広い枠組で議論されている（von Mises [22], Filippova [10], Fernholz [9], Denker [7], Rubin and Vitale [19], Dynkin and Mandelbaum [8], Mandelbaum and Taqqu [14] 等参照）.

仮説が未知母数を含む場合に estimated empirical process を考えれば，この確率過程は或るガウス過程に法則収束する．ところが，この極限過程は，仮説の分布型，用いた推定量，真の母数に依存する．このため複合仮説の適合度検定が困難

であったが，理論的には，そのマルチンゲール項を考えることによって distribution-free な方法の可能性が示された（Khmaladze [13]）．2節では，単純仮説検定問題に関して，経験過程のマルチンゲール項に基づく具体的な統計量の性質を議論する．3節では 対称性の検定に関して基本的な或る極限定理について述べる．4節では経験過程及び経験過程のマルチンゲール項の L_1-ノルムに基づいた検定統計量について述べる．

2．経験分布関数のマルチンゲール項に基づいた統計量

$U_1, U_2, \cdots, U_n$ は $(0,1)$ 上の一様分布に従う独立確率変数列とする．$\Gamma_n(t)$ を $U_1, \cdots, U_n$ から作られる経験分布関数とし，

$$W_n(t) = \sqrt{n}\left(\Gamma_n(t) - \int_0^t \frac{1-\Gamma_n(s)}{1-s}\,ds\right), \quad 0 \le t \le 1$$

と置く．$\{n\Gamma_n(t);\ 0 \leqq t < \infty\}$ を確率過程と見る時，或る filtration に関して canonical counting process になり，その integrated intensity は $n\Lambda_t$ で与えられる．

ここで
$$\Lambda_t = \begin{cases} \int_0^t \frac{1-\Gamma_n(s)}{1-s}\,ds & \text{if } t \leqq 1, \\ \int_0^1 \frac{1-\Gamma_n(s)}{1-s}\,ds & \text{if } t > 1, \end{cases}$$

である．このことから $W_n(t)$ がマルチンゲールになることがわかり，マルチンゲールに対する中心極限定理（例えば Rebolledo [16]）を適用することによって次の定理を導くことができる．

定理2.1. $W_n(t)$ は $D[0,1]$ において，標準Wiener過程へ弱収束する．□

この定理を用いて，より一般的な次の定理が得られる．

定理2.2. F を $[0,1]$ 上の分布関数とし，$X_1,\cdots,X_n$ を F に従う独立確率変数列とする．$F_n(t)$ を経験分布関数とし，

$$Y_n(t)=\sqrt{n}\left(F_n(t)-\int_0^t\frac{1-F_n(s)}{1-F(s)}dF(s)\right)$$

と定義すれば，$Y_n(t)$ は $D[0,1]$ において $W(F(t))$ へ弱収束する．□
ここで $\{W(t):0\le t\le 1\}$ は標準Wiener過程．
（証明略）

さて，F を $(0,1)$ 上の連続な分布関数として，仮説 $F(t)=t$ を検定することを考える．統計量として

$$T_n^S=\sqrt{n}\sup_{0\le t\le 1}\left|F_n(t)-\int_0^t\frac{1-F_n(s)}{1-s}ds\right|$$

をとると，定理2.1と，$\sup|\cdot|$ が $C[0,1]$ 上で連続な事と，Wiener測度が $C[0,1]$ 上に support を持つことから，T_n^S が仮説の下で $\sup_{0\le t\le 1}|W(t)|$ へ法則収束することがわかる．

次に T_n^S を他の統計量と比較することを考える．

定義2.1.（Bahadur [3]）$\{P_\theta:\theta\in\Theta\}$ を可測空間 $(S,\mathcal{S})$ 上の確率測度の族とする．$\Theta_0\subset\Theta$ とし，仮説 $\theta\in\Theta_0$ を H と書く．今，$(S,\mathcal{S})$ 上で定義された統計量の列 $\{T_n\}$（n は sample size）が与えられているとする．このとき $\{T_n\}$ が standard sequence であるとは，次の(I)～(III)が満たされることを言う。

(I) $\exists$ continuous distribution function G s.t.

$\forall \theta \in \Theta_0$, $\lim_{n\to\infty} P(T_n < x) = G(x)$ for every x.

(II) $0 < {}^{\exists}a < \infty$ s.t.

$$\log(1 - G(x)) = -\frac{ax^2}{2}[1 + o(1)],$$

ただし， $o(1) \to 0$ as $x \to \infty$.

(III) ${}^{\exists}b : \Theta \setminus \Theta_0 \to \mathbb{R}$ s.t.

$0 < b(\theta) < \infty$ and

$$\lim_{n\to\infty} P\left(\left|\frac{T_n}{\sqrt{n}} - b(\theta)\right| > x\right) = 0 \quad \text{for every } x > 0.$$

standard sequence $\{T_n\}$ に対して $c(\theta) = a\,b^2(\theta)$ を approximate Bahadur slope という．また，２つの standard sequences $\{T_n^{(1)}\}$ と $\{T_n^{(2)}\}$ について，$c_1(\theta)/c_2(\theta)$ を $\{T_n^{(1)}\}$ の $\{T_n^{(2)}\}$ に対する approximate Bahadur efficiency という．

今，我々の考えている問題では，$S = [0, 1]$，$\{P_\theta, \theta \in \Theta\}$ は $[0, 1]$ 上の分布から成る族である．P_θ の分布関数を F_θ と書く．さらに $\Theta_0 = \{\theta_0\}$ （１点）であり，P_{θ_0} は $[0, 1]$ 上の一様分布である．

さて $[0, 1]$ 上の分布関数 F に関する次の２つの条件を考える．

条件 A．$Y_1, \cdots, Y_n$ を分布関数 $F(1 - e^{-z})$ $(z > 0)$ に従う独立確率変数列とする時，２つの数列 $\{a_n\}_{n=1}^{\infty}$，$\{b_n\}_{n=1}^{\infty}$ と或る分布 Δ が存在して，$n \to \infty$ のとき，$\sqrt{n}\,a_n \to \infty$，$b_n/(\sqrt{n}\,a_n) \to 0$，$a_n \cdot \max\{Y_1, \cdots, Y_n\} + b_n \xrightarrow{\mathcal{D}} \Delta$ が成り立つ．

条件 B. $$\int_0^1 \frac{1-F(t)}{1-t}\,dt < \infty .$$

定理2.3. 任意の $\theta \in \Theta$ に対して, F_θ は条件A, B を満たすとする. このとき統計量の列 $\{T_n^S\}$ は standard sequence になる.

(証明 略)

系. 任意の $\theta \in \Theta$ に対して, F_θ が条件A, B を満たすならば, $\{T_n^S\}$ の approximate Bahadur slope は,

$$c^S(\theta) = \left\{ \sup_{0 \le t \le 1} \left| F_\theta(t) - \int_0^t \frac{1-F_\theta(s)}{1-s}\,ds \right| \right\}^2$$

で与えられる.

次に, 同じ検定問題に対して, 統計量

$$T_n^Q = n \int_0^1 \left\{ F_n(t) - \int_0^t \frac{1-F_n(s)}{1-s}\,ds \right\}^2 dt$$

を考える. 帰無仮説の下では, T_n^Q が $\int_0^1 W^2(t)\,dt$ へ法則収束することが, 定理2.1からわかる.

定理2.4. 任意の $\theta \in \Theta$ に対して, F_θ が条件A, B を満たすならば, $\{(T_n^Q)^{1/2}\}$ は standard sequence になる.

(証明 略)

系. 任意の $\theta \in \Theta$ に対して, F_θ が条件A, B を満たすならば, $\{(T_n^Q)^{1/2}\}$ の approximate Bahadur slope は

$$c^Q(\theta) = \frac{\pi^2}{4} \int_0^1 \left(F_\theta(t) - \int_0^t \frac{1-F_\theta(s)}{1-s}\,ds \right)^2 dt$$

で与えられる.

この節の詳細については Aki[1] を参照.

3. 対称性の検定統計量の一般化

前節で述べた“経験過程のマルチンゲール項”に基づく統計量以外に，その漸近分布がWiener過程の簡単な関数で与えられるものとしては，対称性の検定統計量がよく知られている（Butler [6]，Rothman and Woodroofe [18] 等）．この節では，これらの統計量の漸近分布を導出することに対して基本的な極限定理を少し一般化した形で述べ，対称性の仮説を一般化しても，それが使えることを注意する．

G_1 を $[0,1]$ 上の分布関数とする．$Y_1,\cdots,Y_n$ は独立に G_1 に従う確率変数列，$\xi_1,\cdots,\xi_n$ は独立同分布確率変数列で，任意の Y_i とも独立であるとする．さらに $E\xi_1=0$，$E\xi_1^2=1$ を仮定する．$D[0,1]$ 上の random element $u_n(t)$ を

$$u_n(t)=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i\, I_{[0,t]}(Y_i)$$

と定義する．

定理3.1．上の仮定の下で $u_n(t)$ は $D[0,1]$ において $W(G_1(t))$ へ弱収束する．

（証明の概略）

先ず $G_1(t)=t$ の場合を考える．任意の実数 $a_1,\cdots,a_m$，$0\le t_1,\cdots,t_m\le 1$ に対して，中心極限定理を適用して，

$$\sum_{j=1}^{m}a_j\,u_n(t_j)\xrightarrow{\mathcal{D}}\sum_{j=1}^{m}a_j\,W(t_j)$$

を示すことができ，Cramer－Wold の定理によって有限次元

分布の収束がわかる。tightness に関しては，$0 \le t_1 \le t \le t_2 \le 1$ に対して，u_n の定義から直接，

$$E\{|u_n(t)-u_n(t_1)|^2 \cdot |u_n(t_2)-u_n(t)|^2\} < (t_2-t_1)^2$$

を示すことができるので，Billingsley [5] の Theorem 15.6 によって定理は成立する。

次に $G_1(t)$ が任意の場合を考える。$\varphi(s) \equiv \inf\{t : s \le G_1(t)\}$ によって G_1 の逆関数を定義する。今，$\{\xi_i\}$ と独立に $(0,1)$ 上の一様分布に従う確率変数列 $\eta_1,\cdots,\eta_n$ を用意すると，$\varphi(\eta_i)$ は G_1 に従う。この定理は法則収束に関するものであるから $Y_i = \varphi(\eta_i)$ と書いてもよい。$V_n(t) \equiv \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\xi_i I_{[0,t]}(\eta_i)$ と定義すると $G_1(t)=t$ の場合の証明より，$V_n(t) \xrightarrow{\mathcal{D}} W(t)$ in $D[0,1]$ が成り立っている。また，$u_n(t) = V_n(G_1(t))$ と書け，G_1 による時間変更を $D[0,1]$ 上の写像と見た場合，$C[0,1]$ 上で連続になることから，$u_n(t) \xrightarrow{\mathcal{D}} W(G_1(t))$ in $D[0,1]$ が成り立つ。□

さらに Bickel and Wichura [4] の結果を用いて，u_n の sequential process に関する次の定理を導くことができる。

$Y_1,\cdots,Y_n$ を独立に $(0,1)$ 上の一様分布に従う確率変数列とし，$\xi_1,\cdots,\xi_n$ は定理 3.1 の仮定を満たすとする。$D([0,1]^2)$ 上の random element $u_n(s,t)$ を

$$u_n(s,t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{[ns]}\xi_i I_{[0,t]}(Y_i) \quad , \quad 0 \le s,t \le 1 ,$$

と定義する.

定理 3.2. 上の仮定の下で $u_n(s,t)$ は $D([0,1]^2)$ において Brownian sheet $B(s,t)$ へ弱収束する. ここで $B(s,t)$ は平均 0, $EB(s_1,t_1)B(s_2,t_2)=(s_1\wedge s_2)\cdot(t_1\wedge t_2)$ を満たす $[0,1]^2$ 上の Gaussian random field である.

(証明略)

さて, 次のような仮説検定問題を考える. F を $[0,1]$ 上の分布関数とする. $X_1,\cdots,X_n$ は独立に F に従う確率変数列とし, $0<\alpha<1$ を与えられた定数とする. 今, $X_1,\cdots,X_n$ に基づいて F に関する次の仮説を検定したい.

仮説 H: $[0,1]$ 上の或る分布関数 G が存在して,

$$F(t)=\begin{cases}\alpha G(2t), & 0\le t\le \frac{1}{2},\\ \alpha+(1-\alpha)(1-G(2-2t)), & \frac{1}{2}<t\le 1,\end{cases}$$

と書ける.

この仮説は $t=\frac{1}{2}$ を境にして右と左で分布の形が(定数倍を除いて)等しいということを意味しており, 対称性の仮説の自然な拡張になっている. $i=1,\cdots,n$ に対して,

$$Y_i=\begin{cases}2X_i & \text{if } X_i\le\frac{1}{2},\\ 2(1-X_i) & \text{if } X_i>\frac{1}{2},\end{cases}$$

$$\xi_i=\begin{cases}\sqrt{\frac{1-\alpha}{\alpha}} & \text{if } X_i\le\frac{1}{2},\\ -\sqrt{\frac{\alpha}{1-\alpha}} & \text{if } X_i>\frac{1}{2},\end{cases}$$

と定義すると，仮説Hの下では，$P(Y_i \leq t) = G(t)$ 及び
定理3.1の仮定がすべて満たされる．従って

$$u_n(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_i I_{[0,t]}(Y_i)$$

は仮説Hの下で $W(G(t))$ へ法則収束する．ここで更にGが連続関数であることを仮定すると，例えば統計量
$T_n = \sup_{0 \leq t \leq 1} |u_n(t)|$ によってHを検定できる．T_nは仮説Hの下で $\sup_{0 \leq t \leq 1} |W(t)|$ へ法則収束する．また，この検定は次の意味で一致性を持つ．

定理3.3．Fは $[0,1]$ 上の連続な分布関数で，仮説Hを満たさないと仮定する．このとき，任意の $c > 0$ に対して，
$\lim_{n \to \infty} P(T_n > c) = 1$ が成り立つ．

（証明略）

この節の詳細については Aki[2] を参照．

4．ガウス過程の L_1－ノルムと統計量

Shepp[20]，Rice[17]，Johnson and Killeen[11] によって Brownian bridge の L_1－ノルムの分布が明らかになった．これらの結果により，$T_n = \int_0^1 |\sqrt{n}(F_n(x) - F(x))| dF(x)$ という形の適合度検定統計量が使える．T_nの帰無仮説の下での漸近分布は，$\int_0^1 |B(t)| dt$ の分布である．ここで $\{B(t) : 0 \leq t \leq 1\}$ は Brownian bridge．
以上の事の簡単な紹介が Shorack and Wellner[21] にある．

ところで T_n の代わりに，2節や3節で述べた processes を用いて，$\int_0^1 |W_n(t)|dt$ や $\int_0^1 |u_n(t)|dH_n(t)$ を考えてみる．ここで $H_n(\cdot)$ は $Y_1, \cdots, Y_n$ によって作られる経験分布関数である．これらの場合には，仮説の下での漸近分布は $\int_0^1 |W(t)|dt$ の分布になる．この分布は Kac [12] によって調べられ，その Laplace 変換が知られている．

さて，このような統計量について 2節で紹介した approximate Bahadur slope を計算するのに便利な次の結果を述べる．

定理 4.1. $X(t)$ は $C[0,1]$ 上のガウス過程であるとする．任意の $s, t \in [0,1]$ に対して covariance function $R(s,t) \geq 0$ ならば，
$$\lim_{t\to\infty} \frac{1}{t^2} \log P\left\{\int_0^1 |X(t)|dt > t\right\} = -\frac{1}{2\int_0^1\int_0^1 R(s,t)\,ds\,dt}$$
が成り立つ．

（証明）

証明のアイデアは Marcus and Shepp [15] による．Φ を有理数を端点に持つ区間上で 1 または -1 の値をとる step functions の全体とすれば，
$$\int_0^1 |X(t)|dt = \sup_{\varphi\in\Phi} \int_0^1 X(t)\varphi(t)dt$$
が成り立つ．今，$[0,1]$ の有理数による分割 $0=r_0<r_1<\cdots<r_n=1$ を1つ固定し，φ をこの区間上で ε_i $(\varepsilon_i=\pm1)$ の値をとる step function とする． $Y_i = \int_{r_{i-1}}^{r_i} X(t)dt$ と置けば，
$$X_\varphi = \int_0^1 X(t)\varphi(t)dt = \sum_{i=1}^n \varepsilon_i Y_i.$$

$1 \leq i \leq j \leq n$ に対して，仮定より

$$\mathrm{cov}(Y_i, Y_j) = \int_{r_{i-1}}^{r_i} \left(\int_{r_{j-1}}^{r_j} R(s,t)\, dt \right) ds \geq 0$$

となることに注意すれば

$$\sigma_\varphi^2 = \mathrm{var}\, X_\varphi = \sum_{i=1}^{n} \varepsilon_i^2\, \mathrm{var}(Y_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \varepsilon_i \varepsilon_j\, \mathrm{cov}(Y_i, Y_j)$$

を最大にする ε_i は明らかに $\varepsilon_i = 1$ for all i (又は $\varepsilon_i = -1$ for all i) であり，このとき，

$$\mathrm{var}\, X_\varphi = \mathrm{var} \int_0^1 X(t)\, dt = \int_0^1 \int_0^1 R(s,t)\, ds\, dt$$

となる．これは $[0,1]$ の分割の仕方に依らないから，

$$\sigma^2 = \sup_{\varphi \in \Phi} \mathrm{var}\, X_\varphi = \int_0^1 \int_0^1 R(s,t)\, ds\, dt$$

となる。従って Marcus and Shepp [15] の Theorem 2.5 により，定理は成り立つ．□

特に $X(t) = W(t)$ ならば $\sigma^2 = \frac{1}{3}$，

$X(t) = \rho(t)$ ならば $\sigma^2 = \frac{1}{12}$ として定理が使える。

References

[1] Aki, S. (1986). Some test statistics based on the martingale term of the empirical distribution function, *Ann. Inst. Statist. Math.*, **38**, 1-21.

[2] Aki, S. (1986). On nonparametric tests for symmetry, *Research Memorandum No. 304*, The Institute of Statistical Mathematics.

[3] Bahadur, R. R. (1960). Stochastic comparison of tests, *Ann. Math. Statist.*, **31**, 276-295.

[4] Bickel, P. J. and Wichura, M. J. (1971). Convergence criteria for multiparameter stochastic process and some applications, *Ann. Math. Statist.*, **42**, 1656-1670.

[5] Billingsley, P. (1968). *Convergence of Probability Measures*, Wiley, New York.

[6] Butler, C. (1969). A test for symmetry using the sample distribution function, *Ann. Math. Statist.*, **40**, 2209-2210.
[7] Denker, M. (1985). *Asymptotic Distribution Theory in Nonparametric Statistics*, Vieweg.
[8] Dynkin, E. B. and Mandelbaum, A. (1983). Symmetric statistics, Poisson point processes, and multiple Wiener integrals, *Ann. Statist.*, **11**, 739-745.
[9] Fernholz, L. T. (1983). *von Mises Calculus for Statistical Functionals*, Lecture Notes in Statistics 19, Springer.
[10] Filippova, A. A. (1962). Mises' theorem of the asymptotic behavior of functionals of empirical distribution functions and its statistical applications, *Theory Prob. Appl.*, **7**, 24-57.
[11] Johnson, B. McK. and Killeen, T. (1983). An explicit formula for the cdf of the L_1 norm of the Brownian bridge, *Ann. Prob.*, **11**, 807-808.
[12] Kac, M. (1946). On the average of a certain Wiener functional and a related limit theorem in calculus of probability, *Trans. Amer. Math. Soc.*, **59**, 401-414.
[13] Khmaladze, E. V. (1981). Martingale approach in the theory of goodness-of-fit tests, *Theory Prob. Appl.*, **26**, 240-257.
[14] Mandelbaum, A. and Taqq, M. S. (1984). Invariance principle for symmetric statistics, *Ann. Statist.*, **12**, 483-496.
[15] Marcus, M. B. and Shepp, L. A. (1971). Sample behavior of Gaussian processes, *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, **2**, 423-442.
[16] Rebolledo, R. (1978). Sur les applications de la théorie des martingales à l'étude statistique d'une famille de processus ponctuels, Lecture Notes in Mathematics, Vol. 636, 27-70, Springer, New York.
[17] Rice, S. O. (1982). The integral of the absolute value of the pinned Wiener process-calculation of its probability density by numerical integration, *Ann. Prob.*, **10**, 240-243.
[18] Rothman, E. D. and Woodroofe, M. (1972). A Cramér-von Mises type statistics for testing symmetry, *Ann. Math. Statist.*, **43**, 2035-2038.
[19] Rubin, H. and Vitale, R. A. (1980). Asymptotic distribution of symmetric statistics, *Ann. Statist.*, **8**, 165-170.
[20] Shepp, L. A. (1982). On the integral of the absolute value of the pinned Wiener process, *Ann. Prob.*, **10**, 234-239.
[21] Shorack, G. R. and Wellner, J. A. (1986). *Empirical Processes with Applications to Statistics*, John Wiley & Sons, New York.
[22] von Mises, R. (1947). On the asymptotic distribution of differentiable statistical functions, *Ann. Math. Statist.*, **18**, 309-348.